Sequential Quantile Prediction of Time Series

Gérard Biau, Benoît Patra

Sequential Quantile Prediction of Time Series

Gérard BIAU [a,*] and Benoît PATRA [a,b,†]

[a] LSTA
Université Pierre et Marie Curie – Paris VI
Boîte 158, 175 rue du Chevaleret
75013 Paris, France
gerard.biau@upmc.fr

[b] LOKAD SAS
70 rue Lemercier
75017 Paris, France
benoit.patra@lokad.com

Abstract

Motivated by a broad range of potential applications, we address the quantile prediction problem of real-valued time series. We present a sequential quantile forecasting model based on the combination of a set of elementary nearest neighbor-type predictors called "experts" and show its consistency under a minimum of conditions. Our approach builds on the methodology developed in recent years for prediction of individual sequences and exploits the quantile structure as a minimizer of the so-called pinball loss function. We perform an in-depth analysis of real-world data sets and show that this nonparametric strategy generally outperforms standard quantile prediction methods.

Index Terms — Time series, quantile prediction, pinball loss, sequential prediction, nearest neighbor estimation, consistency, expert aggregation.

AMS 2000 Classification: 62G08; 62G05; 62G20.

*Partially supported by the French "Agence Nationale pour la Recherche" under grant ANR-09-BLAN-0051-02 "CLARA". Research carried out within the INRIA project "CLASSIC" hosted by Ecole Normale Supérieure and CNRS.

†Corresponding author.

1 Introduction

Forecasting the future values of an observed time series is an important problem, which has been an area of considerable activity in recent years. The application scope is vast, as time series prediction applies to many fields, including problems in genetics, medical diagnoses, air pollution forecasting, machine condition monitoring, financial investments, production planning, sales forecasting and stock controls.

To fix the mathematical context, suppose that at each time instant $n = 1, 2, \ldots$, the forecaster (also called the predictor hereafter) is asked to guess the next outcome y_n of a sequence of real numbers $y_1, y_2, \ldots$ with knowledge of the past $y_1^{n-1} = (y_1, \ldots, y_{n-1})$ (where y_1^0 denotes the empty string). Formally, the strategy of the predictor is a sequence $g = \{g_n\}_{n=1}^{\infty}$ of forecasting functions

$$g_n : \mathbb{R}^{n-1} \to \mathbb{R}$$

and the prediction formed at time n is just $g_n(y_1^{n-1})$. Throughout the paper we will suppose that y_1, $y_2, \ldots$ are realizations of random variables Y_1, $Y_2, \ldots$ such that the stochastic process $\{Y_n\}_{-\infty}^{\infty}$ is jointly stationary and ergodic.

Many of the statistical techniques used in time series prediction are those of regression analysis, such as classical least square theory, or are adaptations or analogues of them. These forecasting schemes are typically concerned with finding a function g_n such that the prediction $g_n(Y_1^{n-1})$ corresponds to the conditional mean of Y_n given the past sequence Y_1^{n-1}, or closely related quantities. Many methods have been developed for this purpose, ranging from parametric approaches such as AR(p) and ARMA(p,q) processes (Brockwell and Davies [1]) to more involved nonparametric methods (see for example Györfi et al. [2] and Bosq [3] for a review and references).

On the other hand, while these estimates of the conditional mean serve their purpose, there exists a large area of problems where the forecaster is more interested in estimating conditional quantiles and prediction intervals, in order to know other features of the conditional distribution. There is now a fast pace growing literature on quantile regression (see Gannoun, Saracco and Yu [4] for an overview and references) and considerable practical experience with forecasting methods based on this theory. Economics makes a persuasive case for the value of going beyond models for the conditional mean (Koenker and Allock [5]). In financial mathematics and financial risk management, quantile regression is intimately linked to the τ-Value at Risk

(VaR), which is defined as the $(1-\tau)$-quantile of the portfolio. For example, if a portfolio of stocks has a one-day 5%-VaR of €1 million, there is a 5% probability that the portfolio will fall in value by more than €1 million over a one day period (Duffie and Pan [6]). More generally, quantile regression methods have been deployed in social sciences, ecology, medicine and manufacturing process management. For a description, practical guide and extensive list of references on these methods and related methodologies, we refer the reader to the monograph of Koenker [7].

Motivated by this broad range of potential applications, we address in this paper the quantile prediction problem of real-valued time series. Our approach is nonparametric in spirit and breaks with at least three aspects of more traditional procedures. First, we do not require the series to necessarily satisfy the classical statistical assumptions for bounded, autoregressive or Markovian processes. Indeed, our goal is to show powerful consistency results under a strict minimum of conditions. Secondly, building on the methodology developed in recent years for prediction of individual sequences, we present a sequential quantile forecasting model based on the combination of a set of elementary nearest neighbor-type predictors called "experts". The paradigm of prediction with expert advice was first introduced in the theory of machine learning as a model of online learning in the 1980-early 1990s, and it has been extensively investigated ever since (see the monograph of Cesa-Bianchi and Lugosi [8] for a comprehensive introduction to the domain). Finally, in opposition to standard nonparametric approaches, we attack the problem by fully exploiting the quantile structure as a minimizer of the so-called pinball loss function (Koenker and Basset [9]).

The document is organized as follows. After some basic recalls in Section 2, we present in Section 3 our expert-based quantile prediction procedure and state its consistency under a minimum of conditions. We perform an in-depth analysis of real-world data sets and show that the nonparametric strategy we propose is faster and generally outperforms traditional methods in terms of average prediction errors (Section 4). Proofs of the results are postponed to Section 5.

2 Consistent quantile prediction

2.1 Notation and basic definitions

Let Y be a real-valued random variable with distribution function F_Y, and let $\tau \in (0,1)$. Recall that the generalized inverse of F_Y

$$F_Y^{\leftarrow}(\tau) = \inf\{t \in \mathbb{R} : F_Y(t) \geq \tau\}$$

is called the quantile function of F_Y and that the real number $q_\tau = F_Y^{\leftarrow}(\tau)$ defines the τ-quantile of F_Y (or Y). The basic strategy behind quantile estimation arises from the observation that minimizing the ℓ_1-loss function yields the median. Koenker and Basset [9] generalized this idea and characterized the τ-quantile by tilting the absolute value function in a suitable fashion.

Lemma 2.1 *Let Y be an integrable real-valued random variable and, for $\tau \in (0,1)$, let the map*

$$\rho_\tau(y) = y(\tau - \mathbf{1}_{[y \leq 0]}).$$

Then the quantile q_τ satisfies the property

$$q_\tau \in \operatorname*{argmin}_{q \in \mathbb{R}} \mathbb{E}\left[\rho_\tau(Y-q)\right]. \tag{2.1}$$

Moreover, if F_Y is (strictly) increasing, then the minimum is unique, that is

$$\{q_\tau\} = \operatorname*{argmin}_{q \in \mathbb{R}} \mathbb{E}\left[\rho_\tau(Y-q)\right].$$

We have not been able to find a complete proof of this result, and we briefly state it in Section 5. The function ρ_τ, shown in Figure 2.1, is called the pinball function. For example, for $\tau = 1/2$, it yields back the absolute value function and, in this case, Lemma 2.1 just expresses the fact that the median $q_{1/2} = F^{\leftarrow}(1/2)$ is a solution of the minimization problem

$$q_{1/2} \in \operatorname*{argmin}_{q \in \mathbb{R}} \mathbb{E}|Y-q|.$$

These definitions may be readily extended to pairs (X,Y) of random variables with conditional distribution $F_{Y|X}$. In this case, the conditional quantile $q_\tau(X)$ is the measurable function of X almost surely (a.s.) defined by

$$q_\tau(X) = F_{Y|X}^{\leftarrow}(\tau) = \inf\{t \in \mathbb{R} : F_{Y|X}(t) \geq \tau\},$$

and, as in Lemma 2.1, it can be shown that for an integrable Y

$$q_\tau(X) \in \operatorname*{argmin}_{q(.)} \mathbb{E}_{\mathbb{P}_{Y|X}}\left[\rho_\tau\left(Y - q(X)\right)\right], \tag{2.2}$$

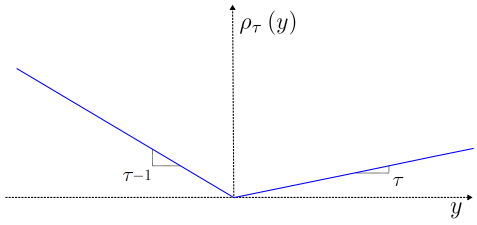

Figure 1: Pinball loss function ρ_τ.

where the infimum is taken over the set of all measurable real-valued functions and the notation $\mathbb{E}_{\mathbb{P}_{Y|X}}$ stands for the conditional expectation of Y with respect to X. We note again that if $F_{Y|X}$ is a.s. increasing, then the solution of (2.2) is unique and equals $q_\tau(X)$ a.s. In the sequel, we will denote by $\mathcal{Q}_\tau(\mathbb{P}_{Y|X})$ the set of solutions of the minimization problem (2.2), so that $q_\tau(X) \in \mathcal{Q}_\tau(\mathbb{P}_{Y|X})$ and $\{q_\tau(X)\} = \mathcal{Q}_\tau(\mathbb{P}_{Y|X})$ when the minimum is unique.

2.2 Quantile prediction

In our sequential version of the quantile prediction problem, the forecaster observes one after another the realizations $y_1, y_2, \ldots$ of a stationary and ergodic random process $Y_1, Y_2, \ldots$ At each time $n = 1, 2, \ldots$, before the n-th value of the sequence is revealed, his mission is to guess the value of the conditional quantile

$$q_\tau(Y_1^{n-1}) = F^{\leftarrow}_{Y_n|Y_1^{n-1}}(\tau) = \inf\{t \in \mathbb{R} : F_{Y_n|Y_1^{n-1}}(t) \geq \tau\},$$

on the basis of the previous $n-1$ observations $Y_1^{n-1} = (Y_1, \ldots, Y_{n-1})$ only. Thus, formally, the strategy of the predictor is a sequence $g = \{g_n\}_{n=1}^{\infty}$ of quantile prediction functions

$$g_n : \mathbb{R}^{n-1} \to \mathbb{R}$$

and the prediction formed at time n is just $g_n(y_1^{n-1})$. After n time instants, the (normalized) cumulative quantile loss on the string Y_1^n is

$$L_n(g) = \frac{1}{n}\sum_{t=1}^{n} \rho_\tau\left(Y_t - g_t(Y_1^{t-1})\right).$$

Ideally, the goal is to make $L_n(g)$ small. There is, however, a fundamental limit for the quantile predictability, which is determined by a result of Algoet

[10]: for any quantile prediction strategy g and jointly stationary ergodic process $\{Y_n\}_{-\infty}^{\infty}$,

$$\liminf_{n\to\infty} L_n(g) \geq L^\star \quad \text{a.s.}, \tag{2.3}$$

where

$$L^\star = \mathbb{E}\left[\min_{q(.)} \mathbb{E}_{\mathbb{P}_{Y_0|Y_{-\infty}^{-1}}}\left[\rho_\tau\left(Y_0 - q(Y_{-\infty}^{-1})\right)\right]\right]$$

is the expected minimal quantile loss over all quantile estimations of Y_0 based on the infinite past observation sequence $Y_{-\infty}^{-1} = (\ldots, Y_{-2}, Y_{-1})$. Generally, we cannot hope to design a strategy whose prediction error exactly achieves the lower bound $L^\star$. Rather, we require that $L_n(g)$ gets arbitrarily close to $L^\star$ as n grows. This gives sense to the following definition:

Definition 2.1 *A quantile prediction strategy g is called consistent with respect to a class $\mathcal{C}$ of stationary and ergodic processes $\{Y_n\}_{-\infty}^{\infty}$ if, for each process in the class,*

$$\lim_{n\to\infty} L_n(g) = L^\star \quad \textit{a.s.}$$

Thus, consistent strategies asymptotically achieve the best possible loss for all processes in the class. In the context of prediction with squared loss, Györfi and Lugosi [11], Nobel [12], Györfi and Ottucsák [13] and Biau et al. [14] study various sequential prediction strategies, and state their consistency under a minimum of assumptions on the collection $\mathcal{C}$ of stationary and ergodic processes. Roughly speaking, these methods consider several "simple" nonparametric estimates (called experts in this context) and combine them at time n according to their past performance. For this, a probability distribution on the set of experts is generated, where a "good" expert has relatively large weight, and the average of all experts' predictions is taken with respect to this distribution. Interestingly, related schemes have been proposed in the context of sequential investment strategies for financial markets. Sequential investment strategies are allowed to use information about the market collected from the past and determine at the beginning of a training period a portfolio, that is, a way to distribute the current capital among the available assets. Here, the goal of the investor is to maximize his wealth in the long run, without knowing the underlying distribution generating the stock prices. For more information on this subject, we refer the reader to Algoet [15], Györfi and Schäfer [16], Györfi, Lugosi and Udina [17], and Györfi, Udina and Walk [18].

Our purpose in this paper will be to investigate an expert-oriented strategy for quantile forecasting. With this aim in mind, we define in the next section

a quantile prediction strategy, called nearest neighbor-based strategy, and state its consistency with respect to a large class of stationary and ergodic processes.

3 A nearest neighbor-based strategy

The quantile prediction strategy is defined at each time instant as a convex combination of elementary predictors (the so-called experts), where the weighting coefficients depend on the past performance of each elementary predictor. To be more precise, we first define an infinite array of experts $h_n^{(k,\ell)}$, where k and ℓ are positive integers. The integer k is the length of the past observation vectors being scanned by the elementary expert and, for each ℓ, choose $p_\ell \in (0,1)$ such that

$$\lim_{\ell\to\infty} p_\ell = 0\,,$$

and set

$$\bar{\ell} = \lfloor p_\ell n \rfloor$$

(where $\lfloor . \rfloor$ is the floor function). At time n, for fixed k and ℓ ($n > k+\bar{\ell}+1$), the expert searches for the $\bar{\ell}$ nearest neighbors (NN) of the last seen observation y_{n-k}^{n-1} in the past and predicts the quantile accordingly. More precisely, let

$$J_n^{(k,\ell)} = \left\{ k < t < n : y_{t-k}^{t-1} \text{ is among the } \bar{\ell}\text{-NN of } y_{n-k}^{n-1} \text{ in } y_1^k, \ldots, y_{n-k-1}^{n-2} \right\},$$

and define the elementary predictor $\bar{h}_n^{(k,\ell)}$ by

$$\bar{h}_n^{(k,\ell)} \in \operatorname*{argmin}_{q\in\mathbb{R}} \sum_{t\in J_n^{(k,\ell)}} \rho_\tau(y_t - q)$$

if $n > k + \bar{\ell} + 1$, and 0 otherwise. Next, let the truncation function

$$T_a(z) = \begin{cases} a & \text{if } z > a; \\ z & \text{if } |z| \le a; \\ -a & \text{if } z < -a, \end{cases}$$

and let

$$h_n^{(k,\ell)} = T_{\min(n^\delta,\ell)} \circ \bar{h}_n^{(k,\ell)}, \tag{3.1}$$

where δ is a positive parameter to be fixed later on. We note that the expert $h_n^{(k,\ell)}$ can be interpreted as a (truncated) $\bar{\ell}$-nearest neighbor regression function estimate drawn in $\mathbb{R}^k$ (Györfi et al. [19]). The proposed quantile prediction algorithm proceeds with an exponential weighting average of the

experts. More formally, let $\{b_{k,\ell}\}$ be a probability distribution on the set of all pairs (k,ℓ) of positive integers such that for all k and ℓ, $b_{k,\ell} > 0$. Fix a learning parameter $\eta_n > 0$, and define the weights

$$w_{k,\ell,n} = b_{k,\ell} e^{-\eta_n (n-1) L_{n-1}(h_n^{(k,\ell)})}$$

and their normalized values

$$p_{k,\ell,n} = \frac{w_{k,\ell,n}}{\sum_{i,j=1}^{\infty} w_{i,j,n}}.$$

The quantile prediction strategy g at time n is defined by

$$g_n(y_1^{n-1}) = \sum_{k,\ell=1}^{\infty} p_{k,\ell,n} h_n^{(k,\ell)}(y_1^{n-1}), \qquad n = 1, 2, \ldots \tag{3.2}$$

The idea of combining a collection of concurrent estimates was originally developed in a non-stochastic context for online sequential prediction from deterministic sequences (Cesa-Bianchi and Lugosi [8]). Following the terminology of the prediction literature, the combination of different procedures is sometimes termed aggregation in the stochastic context. The overall goal is always the same: use aggregation to improve prediction. For a recent review and an updated list of references, see Bunea and Nobel [20].

In order to state consistency of the method, we shall impose the following set of assumptions:

($H1$) One has $\mathbb{E}[Y_0^2] < \infty$.

($H2$) For any vector $\mathbf{s} \in \mathbb{R}^k$, the random variable $\|Y_1^k - \mathbf{s}\|$ has a continuous distribution function.

($H3$) The conditional distribution function $F_{Y_0|Y_{-\infty}^{-1}}$ is a.s. increasing.

Condition ($H2$) expresses the fact that ties occur with probability zero. A discussion on how to deal with ties that may appear in some cases can be found in [21], in the related context of portfolio selection strategies. Condition ($H3$) is mainly technical and ensures that the minimization problem (2.2) has a unique solution or, put differently, that the set $\mathcal{Q}_\tau(\mathbb{P}_{Y_0|Y_{-\infty}^{-1}})$ reduces to the singleton $\{F^{\leftarrow}_{Y_0|Y_{-\infty}^{-1}}(\tau)\}$.

We are now in a position to state the main result of the paper.

Theorem 3.1 *Let $\mathcal{C}$ be the class of all jointly stationary ergodic processes $\{Y_n\}_{-\infty}^{\infty}$ satisfying conditions $(H1)$-$(H3)$. Suppose in addition that $n\eta_n \to \infty$ and $n^{2\delta}\eta_n \to 0$ as $n \to \infty$. Then the nearest neighbor quantile prediction strategy defined above is consistent with respect to $\mathcal{C}$.*

The truncation index T in definition (3.1) of the elementary expert $h_n^{(k,\ell)}$ is merely a technical choice that avoids having to assume that $|Y_0|$ is a.s. bounded. On the practical side, it has little influence on results for relatively short time series. On the other hand, the choice of the learning parameter η_n as $1/\sqrt{n}$ ensures consistency of the method for $0 < \delta < \frac{1}{4}$.

4 Experimental results

4.1 Algorithmic settings

In this section, we evaluate the behavior of the nearest neighbor quantile prediction strategy on real-world data sets and compare its performances to those of standard families of methods on the same data sets.

Before testing the different procedures, some precisions on the computational aspects of the presented method are in order. We first note that infinite sums make formula (3.2) impracticable. Thus, for practical reasons, we chose a finite grid $(k, \ell) \in \mathcal{K} \times \mathcal{L}$ of experts (positive integers), let

$$g_n(y_1^{n-1}) = \sum_{k\in\mathcal{K},\ell\in\mathcal{L}} p_{k,\ell,n} h_n^{(k,\ell)}(y_1^{n-1}), \qquad n = 1, 2, \ldots \tag{4.1}$$

and fixed the probability distribution $\{q_{k,\ell}\}$ as the uniform distribution over the $|\mathcal{K}|\times|\mathcal{L}|$ experts. Observing that $h_n^{(k,\ell_1)} = h_n^{(k,\ell_2)}$ and $b_{k,\ell_1} = b_{k,\ell_2}$ whenever $\bar{\ell_1} = \bar{\ell_2}$, formula (4.1) may be more conveniently rewritten as

$$g_n(y_1^{n-1}) = \sum_{k\in\mathcal{K},\bar{\ell}\in\bar{\mathcal{L}}} p_{k,\bar{\ell},n} h_n^{(k,\bar{\ell})}(y_1^{n-1}),$$

where $\bar{\mathcal{L}} = \{\bar{\ell} : \ell \in \mathcal{L}\}$. In all subsequent numerical experiments, we chose $\mathcal{K} = \{1, 2, 3, \ldots, 14\}$ and $\bar{\mathcal{L}} = \{1, 2, 3, \ldots, 25\}$.

Next, as indicated by the theoretical results, we fixed $\eta_n = \sqrt{1/n}$. For a thorough discussion on the best practical choice of η_n, we refer to [14]. To avoid numerical instability problems while computing the $p_{k,\bar{\ell},n}$, we applied if necessary a simple linear transformation on all $L_n(h_n^{(k,\bar{\ell})})$, just to force

these quantities to belong to an interval where the effective computation of $x \mapsto \exp(-x)$ is numerically stable.

Finally, in order to deal with the computation of the elementary experts (3.1), we denote by $\lceil . \rceil$ the ceiling function and observe that if $m \times \tau$ is not an integer, then the solution of the minimization problem $\operatorname{argmin}_{b \in \mathbb{R}} \sum_{i=1}^{m} \rho_\tau (y_i - b)$ is unique and equals the $\lceil m \times \tau \rceil$-th element in the sorted sample list. On the other hand, if $m \times \tau$ is an integer then the minimum is not unique, but the $m \times \tau$-th element in the sorted sequence may be chosen as a minimizer. Thus, practically speaking, each elementary expert is computed by sorting the sample. The complexity of this operation is $\mathrm{O}(\bar{\ell} \log(\bar{\ell}))$—it is almost linear and feasible even for large values of $\bar{\ell}$. For a more involved discussion, we refer the reader to Koenker [7].

All algorithms have been implemented using the oriented object language C# 3.0 and .NET Framework 3.5.

4.2 Data sets and results

We investigated 21 real-world time series representing daily call volumes entering call centers. Optimizing the staff level is one of the most difficult and important tasks for a call center manager. Indeed, if the staff is overdimensioned, then most of the employees will be inactive. On the other hand, underestimating the staff may lead to long waiting phone queues of customers. Thus, in order to know the right staff level, the manager needs to forecast the call volume series and, to get a more accurate staff level planning, he has to forecast the quantiles of the series.

In our data set the series had on average 760 points, ranging from 383 for the shortest to 826 for the longest. Four typical series are shown in Figure 4.2.
We used a set $\mathcal{D}$ of selected dates $m < n$ and, for each method and each time series $(y_1, \ldots, y_n)$ (here, $n = 760$ on average), we trained the models on the pruned series $(y_1, \ldots, y_m)$ and predicted the τ-quantile at time $m + 1$. The set $\mathcal{D}$ is composed of 91 dates, so that all quality criteria used to measure the proximity between the predicted quantiles and the observed values y_{m+1} were computed using $91 \times 21 = 1911$ points. The 21 times series and the set $\mathcal{D}$ are available at the address `http://www.lsta.upmc.fr/doct/patra/` .

In a first series of experiments, we let the methods predict the τ-quantiles at the 1911 dates for $\tau \in \{0.1, 0.5, 0.9\}$. We compared the performances of

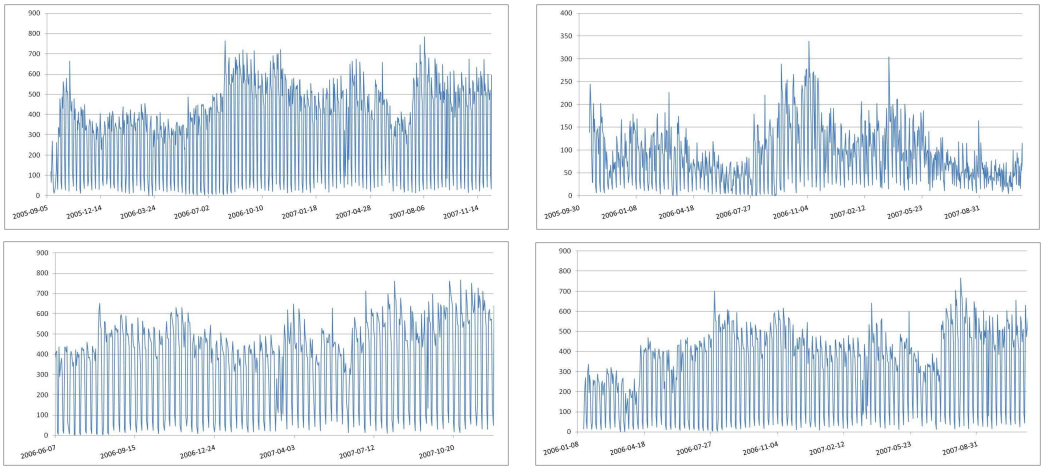

Figure 2: Four call center series, out of 21.

our expert-based strategy, denoted hereafter by $\texttt{QuantileExpertMixture}_\tau$, with those of $\texttt{QAR(p)}_\tau$, a τ-quantile linear autoregressive model of order p. This quantile prediction model, which is described in [7], also uses the pinball criterion to fit its parameters. The implementation we used solves the minimization problem with an Iterative Re-weighted Least Square algorithm (IRLS), see for instance Street, Caroll and Ruppert [22]. Following Takeuchi, Le, Sears and Smola [23], we used two criteria to measure the quality of the overall set of quantile forecastings. First, we evaluated the expected risk with respect to the pinball function ρ_τ, referred to as PinBall Loss in the sequel. Secondly we calculated Ramp Loss, the empirical fraction of quantile estimates which exceed the observed values y_{m+1}. Ideally, the value of Ramp Loss should be close to $1 - \tau$.

Tables 1-3 show the $\texttt{QuantileExpertMixture}_\tau$ and $\texttt{QAR(p)}_\tau$ results at the selected dates $\mathcal{D}$ of the call center series. The latter algorithm was benchmarked for each order p in $\{1, \ldots, 10\}$, but we reported only the most accurate order $p = 7$. The best results with respect to each criterion are shown in bold. We see that both methods perform roughly similarly, with eventually a slight advantage for the autoregressive strategy for $\tau = 0.1$ whereas $\texttt{QuantileExpertMixture}_\tau$ does better for $\tau = 0.9$.

Method	PinBall Loss (0.1)	Ramp Loss
$\texttt{QuantileExpertMixture}_{0.1}$	13.71	0.80
$\texttt{QAR(7)}_{0.1}$	**13.22**	**0.88**

Table 1: Quantile forecastings with $\tau = 0.1$.

Method	PinBall Loss (0.5)	Ramp Loss
QuantileExpertMixture$_{0.5}$	**24.05**	0.42
QAR(7)$_{0.5}$	29.157	**0.47**

Table 2: Quantile forecastings with $\tau = 0.5$.

Method	PinBall Loss (0.9)	Ramp Loss
QuantileExpertMixture$_{0.9}$	**12.27**	**0.07**
QAR(7)$_{0.9}$	19.31	**0.07**

Table 3: Quantile forecastings with $\tau = 0.9$.

Median-based predictors are well known for their robustness while predicting individual values for time series, see for instance Hall, Peng and Yao [24]. Therefore, in a second series of experiments, we fixed $\tau = 0.5$ and focused on the problem of predicting future outcomes of the series. We decided to compare the results of QuantileExpertMixture$_{0.5}$ with those of 6 concurrent predictive procedures:

- MA denotes the simple moving average model.
- AR(p) is a linear autoregressive model of order p, with parameters computed with respect to the usual least square criterion.
- QAR(p) is the τ-quantile linear autoregressive model of order p described earlier.
- DayOfTheWeekMA is a naive model, which applies moving averages on the days of the week, that is a moving average on the Sundays, Mondays, and so on.
- MeanExpertMixture is an online prediction algorithm described in [14]. It is based on conditional mean estimation and close in spirit to the strategy QuantileExpertMixture$_{0.5}$.
- And finally, we let HoltWinters be the well-known procedure which performs exponential smoothing on three components of the series, namely Level, Trend and Seasonality. For a thorough presentation of HoltWinters techniques we refer the reader to Madrikakis, Whellwright and Hyndman [25].

Accuracy of all forecasting methods were measured using the Average Absolute Error (Avg Abs Error, which is proportional to the pinball error since $\tau = 0.5$), Average Squared Error (Avg Sqr Error), and the unstable but widely spread criterion Mean Average Percentage Error (MAPE, see [25] for definition and discussion). We also reported the figure Abs Std Dev which corresponds to the empirical standard deviation of the differences $|y_t^F - y_t^R|$, where y_t^F stands for the

forecasted value while y_t^R stands for the observed value of the time series at time t. `AR(p)` and `QAR(p)` algorithms were run for each order p in $\{1, \ldots, 10\}$, but we reported only the most accurate orders.

Method	Avg Abs Error	Avg Sqr Error	MAPE (%)	Abs Std Dev
`MA`	179.0	62448	52.0	174.8
`AR(7)`	65.8	9738	31.6	73.5
$\texttt{QAR(8)}_{0.5}$	57.8	9594	24.9	79.2
`DayOfTheWeekMA`	54.1	7183	22.8	64.7
$\texttt{QuantileExpertMixture}_{0.5}$	**48.1**	**5731**	21.6	**58.4**
`MeanExpertMixture`	52.4	6536	22.3	61.6
`HoltWinters`	49.8	6025	**21.5**	59.5

Table 4: Future outcomes forecastings.

We see via Table 4 that the nearest neighbor strategy presented here outperforms all other methods in terms of Average Absolute Error. Interestingly, this forecasting procedure also provides the best results with respect to the Average Squared Error criterion. This is remarkable, since $\texttt{QuantileExpertMixture}_{0.5}$ does not rely on a squared error criterion, contrary to `MeanExpertMixture`. The same comment applies to $\texttt{QAR(8)}_{0.5}$ and `AR(7)`. In terms of the Mean Average Percentage Error, the present method and `HoltWinters` procedure provide good and broadly similar results.

5 Proofs

5.1 Proof of Theorem 3.1

The following lemmas will be essential in the proof of Theorem 3.1. The first one is known as Breiman's generalised ergodic theorem (Breiman [26]).

Lemma 5.1 *Let $Z = \{Z_n\}_{-\infty}^{\infty}$ be a stationary and ergodic process. For each positive integer t, let T^t denote the left shift operator, shifting any sequence of real numbers $\{\ldots, z_{-1}, z_0, z_1, \ldots\}$ by t digits to the left. Let $\{f_t\}_{t=1}^{\infty}$ be a sequence of real-valued functions such that $\lim_{t\to\infty} f_t(Z) = f(Z)$ a.s. for some function f. Suppose that $\mathbb{E}[\sup_t |f_t(Z)|] < \infty$. Then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} f_t(T^t Z) = \mathbb{E}\left[f(Z)\right] \quad a.s.$$

Lemma 5.2 below is due to Györfi and Ottucsák [13]. These authors proved the inequality for any cumulative normalized loss of form $L_n(h) = \frac{1}{n}\sum_{t=1}^{n} \ell_t(h)$, where $\ell_t(h) = \ell_t(h_t, Y_t)$ is convex in its first argument, what is the case for the function $\ell_t(h_t, Y_t) = \rho_\tau(Y_t - h_t(Y_1^{t-1}))$.

Lemma 5.2 *Let $g = \{g_n\}_{n=1}^{\infty}$ be the nearest neighbor quantile prediction strategy defined in (3.2). Then, for every $n \geq 1$, a.s.,*

$$L_n(g) \leq \inf_{k,\ell} \left(L_n(h_n^{(k,\ell)}) - \frac{2 \ln b_{k,\ell}}{n\eta_{n+1}} \right) + \frac{1}{2n} \sum_{t=1}^{n} \eta_t \sum_{k,\ell=1}^{\infty} p_{k,\ell,n} \Big[\rho_\tau \left(Y_t - h_t^{(k,\ell)}(Y_1^{t-1}) \right) \Big]^2.$$

Lemma 5.3 *Let $x, y \in \mathbb{R}$ and $\ell \in \mathbb{N}$. Then*

1. $\rho_\tau(x) \leq |x|$.
2. $\rho_\tau(x+y) \leq \rho_\tau(x) + \rho_\tau(y)$.
3. $\rho_\tau \left(T_\ell(x) - T_\ell(y) \right) \leq \rho_\tau(x-y)$.

Proof of Lemma 5.3 Let $x, y \in \mathbb{R}$ and $\ell \in \mathbb{N}$.

1. We have
$$|\rho_\tau(x)| = \left| x(\tau - \mathbf{1}_{[x \leq 0]}) \right| = |x| \left| \tau - \mathbf{1}_{[x \leq 0]} \right| \leq |x|.$$

2. Clearly,
$$\begin{aligned} & \rho_\tau(x+y) \leq \rho_\tau(x) + \rho_\tau(y) \\ \iff \quad & x\mathbf{1}_{[x \leq 0]} + y\mathbf{1}_{[y \leq 0]} \leq x\mathbf{1}_{[x+y \leq 0]} + y\mathbf{1}_{[x+y \leq 0]}. \end{aligned}$$
The conclusion follows by examining the different positions of x and y with respect to 0.

3. If $x > \ell$ and $|y| \leq \ell$, then
$$\begin{aligned} \rho_\tau \left(T_\ell(x) - T_\ell(y) \right) &= \rho_\tau(\ell - y) \\ &= (\ell - y)(\tau - \mathbf{1}_{[\ell - y \leq 0]}) \\ &= (\ell - y)\tau \\ &\leq (x - y)\tau \\ &= (x - y)(\tau - \mathbf{1}_{[x - y \leq 0]}) \\ &= \rho_\tau(x - y). \end{aligned}$$
Similarly, if $x < -\ell$ and $|y| \leq \ell$, then
$$\begin{aligned} \rho_\tau \left(T_\ell(x) - T_\ell(y) \right) &= \rho_\tau(-\ell - y) \\ &= (-\ell - y)(\tau - \mathbf{1}_{[-\ell - y \leq 0]}) \\ &= (-\ell - y)(\tau - 1) \\ &\leq (x - y)(\tau - 1) \\ &= (x - y)(\tau - \mathbf{1}_{[x - y \leq 0]}) \\ &= \rho_\tau(x - y). \end{aligned}$$
All the other cases are similar and left to the reader.

□

Recall that a sequence $\{\mu_n\}_{n=1}^{\infty}$ of probability measures on $\mathbb{R}$ is defined to converge weakly to the probability measure μ_∞ if for every bounded, continuous real function f,

$$\int f \mathrm{d}\mu_n \to \int f \mathrm{d}\mu_\infty \quad \text{as } n \to \infty.$$

Recall also that the sequence $\{\mu_n\}_{n=1}^{\infty}$ is said to be uniformly integrable if

$$\lim_{\alpha \to \infty} \sup_{n \geq 1} \int_{|x| \geq \alpha} |x| \mathrm{d}\mu_n(x) = 0.$$

Moreover, if

$$\sup_{n \geq 1} \int |x|^{1+\varepsilon} \mathrm{d}\mu_n(x) < \infty$$

for some positive ε, then the sequence $\{\mu_n\}_{n=1}^{\infty}$ is uniformly integrable (Billingsley [27]).

The next lemma may be summarized by saying that if a sequence of probability measures converges in terms of weak convergence topology, then the associated quantile sequence will converge too.

Lemma 5.4 *Let $\{\mu_n\}_{n=1}^{\infty}$ be a uniformly integrable sequence of real probability measures, and let μ_∞ be a probability measure with (strictly) increasing distribution function. Suppose that $\{\mu_n\}_{n=1}^{\infty}$ converges weakly to μ_∞. Then, for all $\tau \in (0,1)$,*

$$q_{\tau,n} \to q_{\tau,\infty} \quad \textit{as } n \to \infty,$$

where $q_{\tau,n} \in \mathcal{Q}_\tau(\mu_n)$ for all $n \geq 1$ and $\{q_{\tau,\infty}\} = \mathcal{Q}_\tau(\mu_\infty)$.

Proof of Lemma 5.4 Since $\{\mu_n\}_{n=1}^{\infty}$ converges weakly to μ_∞, it is a tight sequence. Consequently, there is a compact set, say $[-M, M]$, such that $\mu_n(\mathbb{R} \setminus [-M, M]) < \min(\tau, 1-\tau)$. This implies $q_{\tau,n} \in [-M, M]$ for all $n \geq 1$. Consequently, it will be enough to prove that any consistent subsequence of $\{q_{\tau,n}\}_{n=1}^{\infty}$ converges towards $q_{\tau,\infty}$.

Using a slight abuse of notation, we still denote by $\{q_{\tau,n}\}_{n=1}^{\infty}$ a consistent subsequence of the original sequence, and let $q_{\tau,\star}$ be such that $\lim_{n\to\infty} q_{\tau,n} = q_{\tau,\star}$. Using the assumption on the distribution function of μ_∞, we know by Lemma 2.1 that $q_{\tau,\infty}$ is the unique minimizer of problem (2.1). Therefore, to show that $q_{\tau,\star} = q_{\tau,\infty}$, it suffices to prove that, for any $q \in \mathbb{R}$,

$$\mathbb{E}_{\mu_\infty}\left[\rho_\tau(Y - q)\right] \geq \mathbb{E}_{\mu_\infty}\left[\rho_\tau(Y - q_{\tau,\star})\right].$$

Fix $q \in \mathbb{R}$. We first prove that

$$\mathbb{E}_{\mu_n}\left[\rho_\tau(Y - q)\right] \to \mathbb{E}_{\mu_\infty}\left[\rho_\tau(Y - q)\right] \quad \text{as } n \to \infty. \tag{5.1}$$

To see this, for $M > 0$ and all $y \in \mathbb{R}$, set

$$\rho_\tau^{(+,M)}(y) = \begin{cases} 0 & \text{if } |y| < M; \\ \rho_\tau(y) & \text{if } |y| > M+1; \\ \rho_\tau(M+1)(y-M) & \text{if } y \in [M, M+1]; \\ \rho_\tau(-M-1)(y+M) & \text{if } y \in [-M-1, -M]. \end{cases}$$

The function $\rho_\tau^{(+,M)}$ is continuous and, for all $z \in \mathbb{R}$, satisfies the inequality $\rho_\tau^{(+,M)}(z) \leq \rho_\tau(z)\mathbf{1}_{[|z|>M]}$. In the sequel, we will denote by $\rho_\tau^{(-,M)}$ the bounded and continuous map $\rho_\tau - \rho_\tau^{(+,M)}$. The decomposition $\rho_\tau = \rho_\tau^{(+,M)} + \rho_\tau^{(-,M)}$ is illustrated in Figure 5.1.

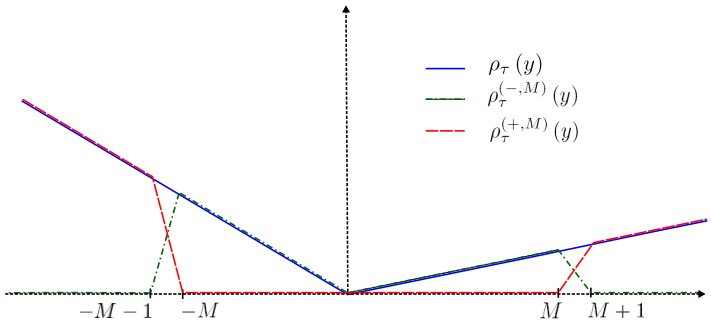

Figure 3: Illustration of the decomposition $\rho_\tau = \rho_\tau^{(+,M)} + \rho_\tau^{(-,M)}$.

Next, fix $\varepsilon > 0$ and choose M large enough to ensure

$$\sup_{n \geq 1} \left(\mathbb{E}_{\mu_n} \left[|Y-q| \mathbf{1}_{[|Y-q|>M]} \right] \right) + \mathbb{E}_{\mu_\infty} \left[|Y-q| \mathbf{1}_{[|Y-q|>M]} \right] < \varepsilon/2.$$

Choose also n sufficiently large to have

$$\left| \mathbb{E}_{\mu_n} \left[\rho_\tau^{(-,M)}(Y-q) \right] - \mathbb{E}_{\mu_\infty} \left[\rho_\tau^{(-,M)}(Y-q) \right] \right| < \varepsilon/2.$$

Write

$$\begin{aligned} &|\mathbb{E}_{\mu_n} [\rho_\tau(Y-q)] - \mathbb{E}_{\mu_\infty} [\rho_\tau(Y-q)]| \\ &\quad = \Big| \mathbb{E}_{\mu_n} \left[\rho_\tau^{(+,M)}(Y-q) \right] + \mathbb{E}_{\mu_n} \left[\rho_\tau^{(-,M)}(Y-q) \right] \\ &\qquad - \mathbb{E}_{\mu_\infty} \left[\rho_\tau^{(+,M)}(Y-q) \right] - \mathbb{E}_{\mu_\infty} \left[\rho_\tau^{(-,M)}(Y-q) \right] \Big|. \end{aligned}$$

Thus

$$
\begin{aligned}
&\left|\mathbb{E}_{\mu_n}\left[\rho_\tau(Y-q)\right]-\mathbb{E}_{\mu_\infty}\left[\rho_\tau(Y-q)\right]\right| \\
&\quad\le \left|\mathbb{E}_{\mu_n}\left[\rho_\tau^{(-,M)}(Y-q)\right]-\mathbb{E}_{\mu_\infty}\left[\rho_\tau^{(-,M)}(Y-q)\right]\right| \\
&\qquad + \left|\mathbb{E}_{\mu_n}\left[\rho_\tau^{(+,M)}(Y-q)\right]-\mathbb{E}_{\mu_\infty}\left[\rho_\tau^{(+,M)}(Y-q)\right]\right| \\
&\quad\le \left|\mathbb{E}_{\mu_n}\left[\rho_\tau^{(-,M)}(Y-q)\right]-\mathbb{E}_{\mu_\infty}\left[\rho_\tau^{(-,M)}(Y-q)\right]\right| \\
&\qquad + \left|\mathbb{E}_{\mu_n}\left[\rho_\tau(Y-q)\mathbf{1}_{[|Y-q|>M]}\right]\right| + \left|\mathbb{E}_{\mu_\infty}\left[\rho_\tau(Y-q)\mathbf{1}_{[|Y-q|>M]}\right]\right| \\
&\quad\le \left|\mathbb{E}_{\mu_n}\left[\rho_\tau^{(-,M)}(Y-q)\right]-\mathbb{E}_{\mu_\infty}\left[\rho_\tau^{(-,M)}(Y-q)\right]\right| \\
&\qquad + \sup_n \mathbb{E}_{\mu_n}\left[|Y-q|\mathbf{1}_{[|Y-q|>M]}\right] + \mathbb{E}_{\mu_\infty}\left[|Y-q|\mathbf{1}_{[|Y-q|>M]}\right] \\
&\quad\le \varepsilon.
\end{aligned}
$$

for all large enough n. This shows (5.1).

Next, using the fact that the function ρ_τ is uniformly continuous, we may write, for sufficiently large n and all $y \in \mathbb{R}$,

$$\rho_\tau(y-q_{\tau,n}) \ge \rho_\tau(y-q_{\tau,\star}) - \varepsilon. \tag{5.2}$$

Therefore, for all large enough n,

$$
\begin{aligned}
\mathbb{E}_{\mu_\infty}\left[\rho_\tau(Y-q)\right] &\ge \mathbb{E}_{\mu_n}\left[\rho_\tau(Y-q)\right]-\varepsilon \\
&\quad \text{(by identity (5.1))} \\
&\ge \mathbb{E}_{\mu_n}\left[\rho_\tau(Y-q_{\tau,n})\right]-\varepsilon \\
&\ge \mathbb{E}_{\mu_n}\left[\rho_\tau(Y-q_{\tau,\star})\right]-2\varepsilon \\
&\quad \text{(by inequality(5.2))} \\
&\ge \mathbb{E}_{\mu_\infty}\left[\rho_\tau(Y-q_{\tau,\star})\right]-3\varepsilon \\
&\quad \text{(by identity (5.1)).}
\end{aligned}
$$

Letting $\varepsilon \to 0$ leads to the desired result.

□

We are now in a position to prove Theorem 3.1.

Because of inequality (2.3) it is enough to show that

$$\limsup_{n\to\infty} L_n(g) \le L^\star \quad \text{a.s.}$$

With this in mind, we first provide an upper bound on the first term of the right hand side of the inequality in Lemma 5.2. We have

$$\begin{aligned}\limsup_{n\to\infty}\inf_{k,\ell}\left(L_n\left(h_n^{(k,\ell)}-\frac{2\ln b_{k,\ell}}{n\eta_{n+1}}\right)\right)&\\ \le\inf_{k,\ell}\left(\limsup_{n\to\infty}L_n\left(h_n^{(k,\ell)}-\frac{2\ln b_{k,\ell}}{n\eta_{n+1}}\right)\right)&\\ \le\inf_{k,\ell}\left(\limsup_{n\to\infty}L_n\left(h_n^{(k,\ell)}\right)\right).&\end{aligned}$$

To evaluate $\limsup_{n\to\infty}L_n(h_n^{(k,\ell)})$, we investigate the performance of the expert $h_n^{(k,\ell)}$ on the stationary and ergodic sequence $Y_0, Y_{-1}, Y_{-2}, \ldots$ Fix $p_\ell \in (0,1)$, $\mathbf{s}\in\mathbb{R}^k$, and set $\tilde{\ell}=\lfloor p_\ell j\rfloor$, where j is a positive integer.

For $j > k+\tilde{\ell}+1$, introduce the set

$$\begin{aligned}\tilde{J}_{j,\mathbf{s}}^{(k,\tilde{\ell})}=\Big\{&-j+k+1\le i\le 0: Y_{i-k}^{i-1}\text{ is among the }\tilde{\ell}\text{-NN of }\mathbf{s}\\ &\text{in } Y_{-k}^{-1},\ldots,Y_{-j+1}^{-j+k}\Big\}.\end{aligned}$$

For any real number a, we denote by δ_a the Dirac (point) measure at a. Let the random measure $\mathbb{P}_{j,\mathbf{s}}^{(k,\ell)}$ be defined by

$$\mathbb{P}_{j,\mathbf{s}}^{(k,\ell)}=\frac{1}{\left|\tilde{J}_{j,\mathbf{s}}^{(k,\tilde{\ell})}\right|}\sum_{i\in\tilde{J}_{j,\mathbf{s}}^{(k,\tilde{\ell})}}\delta_{Y_i}.$$

Take an arbitrary radius $r_{k,\ell}(\mathbf{s})$ such that

$$\mathbb{P}\left[\|Y_{-k}^{-1}-\mathbf{s}\|\le r_{k,\ell}(\mathbf{s})\right]=p_\ell.$$

A straightforward adaptation of an argument in Theorem 3.1 of [18] shows that

$$\mathbb{P}_{j,\mathbf{s}}^{(k,\ell)}\underset{j\to\infty}{\to}\mathbb{P}_{Y_0\,|\,\|Y_{-k}^{-1}-\mathbf{s}\|\le r_{k,\ell}(\mathbf{s})}\triangleq\mathbb{P}_{\infty,\mathbf{s}}^{(k,\ell)}$$

almost surely in terms of weak convergence. Moreover, by a double application of the ergodic theorem (see for instance [14]),

$$\int y^2\mathrm{d}\mathbb{P}_{j,\mathbf{s}}^{(k,\ell)}(y)\underset{j\to\infty}{\to}\int y^2\mathrm{d}\mathbb{P}_{\infty,\mathbf{s}}^{(k,\ell)}(y)\quad\text{a.s.}$$

Thus

$$\sup_{j\ge 0}\int y^2\mathrm{d}\mathbb{P}_{j,\mathbf{s}}^{(k,\ell)}(y)<\infty\quad\text{a.s.},$$

and, consequently, the sequence $\{\mathbb{P}_{j,\mathbf{s}}^{(k,\ell)}\}_{j=1}^{\infty}$ is uniformly integrable.

By assumption $(H3)$ the distribution function of the measure $\mathbb{P}_{Y_0|Y_{-\infty}^{-1}}$ is a.s. increasing. We also have $\sigma\left(\|Y_{-k}^{-1}-\mathbf{s}\| \le r_{k,\ell}(\mathbf{s})\right) \subset \sigma\left(Y_{-\infty}^{-1}\right)$ where $\sigma(X)$ denotes the sigma algebra generated by the random variable X. Thus the distribution function of $\mathbb{P}_{\infty,\mathbf{s}}^{(k,\ell)} = \mathbb{P}_{Y_0\,|\,\|Y_{-k}^{-1}-\mathbf{s}\|\le r_{k,\ell}(\mathbf{s})}$ is a.s. increasing, too. Hence, letting

$$q_{\tau,j}^{(k,\ell)}(Y_{-j+1}^{-1},\mathbf{s}) \in \mathcal{Q}_\tau(\mathbb{P}_{j,\mathbf{s}}^{(k,\ell)}) \quad \text{and} \quad \left\{q_{\tau,\infty}^{(k,\ell)}(\mathbf{s})\right\} = \mathcal{Q}_\tau(\mathbb{P}_{\infty,\mathbf{s}}^{(k,\ell)}),$$

we may apply Lemma 5.4, and obtain

$$q_{\tau,j}^{(k,\ell)}(Y_{-j+1}^{-1},\mathbf{s}) \underset{j\to\infty}{\to} q_{\tau,\infty}^{(k,\ell)}(\mathbf{s}) \qquad \text{a.s.}$$

Consequently, for any $y_0 \in \mathbb{R}$,

$$\rho_\tau\left(y_0 - T_{\min(j^\delta,\ell)}\left(q_{\tau,j}^{(k,\ell)}(Y_{-j+1}^{-1},\mathbf{s})\right)\right) \underset{j\to\infty}{\to} \rho_\tau\left(y_0 - T_\ell\left(q_{\tau,\infty}^{(k,\ell)}(\mathbf{s})\right)\right) \qquad \text{a.s.}$$

Since y_0 and $\mathbf{s}$ are arbitrary, we are led to

$$\rho_\tau\left(Y_0 - T_{\min(j^\delta,\ell)}\left(q_{\tau,j}^{(k,\ell)}(Y_{-j+1}^{-1},Y_{-k}^{-1})\right)\right) \underset{j\to\infty}{\to} \rho_\tau\left(Y_0 - T_\ell\left(q_{\tau,\infty}^{(k,\ell)}(Y_{-k}^{-1})\right)\right) \qquad \text{a.s.} \tag{5.3}$$

For $y = (\ldots, y_{-1}, y_0, y_1, \ldots)$, set

$$\begin{aligned} f_j(y) &\triangleq \rho_\tau\left(y_0 - h_j^{(k,\ell)}(y_{-j+1}^{-1})\right) \\ &= \rho_\tau\left(y_0 - T_{\min(j^\delta,\ell)}\left(q_{\tau,j}^{(k,\ell)}(y_{-j+1}^{-1},y_{-k}^{-1})\right)\right). \end{aligned}$$

Clearly,

$$\begin{aligned} |f_j(Y)| &= \left|\rho_\tau\left(Y_0 - h_j^{(k,\ell)}(Y_{-j+1}^{-1})\right)\right| \\ &\le \left|Y_0 - T_{\min(j^\delta,\ell)}(Y_{-j+1}^{-1})\right| \\ &\quad \text{(by statement 1. of Lemma 5.3)} \\ &\le |Y_0| + \left|T_{\min(j^\delta,\ell)}(Y_{-j+1}^{-1})\right| \\ &\le |Y_0| + \ell, \end{aligned}$$

and thus $\mathbb{E}[\sup_j |f_j(Y)|] < \infty$. By identity (5.3),

$$f_j(Y) \underset{j\to\infty}{\to} \rho_\tau\left(Y_0 - T_\ell(q_{\tau,\infty}^{(k,\ell)}(Y_{-k}^{-1})\right) \quad \text{a.s.}$$

Consequently, Lemma 5.1 yields

$$L_n(h_n^{(k,\ell)}) \underset{n\to\infty}{\to} \mathbb{E}\left[\rho_\tau\left(Y_0 - T_\ell\left(q_{\tau,\infty}^{(k,\ell)}(Y_{-k}^{-1})\right)\right)\right].$$

To lighten notation a bit, we set

$$\varepsilon_{k,\ell} \triangleq \mathbb{E}\left[\rho_\tau\left(Y_0 - T_\ell\left(q_{\tau,\infty}^{(k,\ell)}(Y_{-k}^{-1})\right)\right)\right]$$

and proceed now to prove that $\lim_{k\to\infty}\lim_{\ell\to\infty}\varepsilon_{k,\ell} \leq L^\star$.

We have, a.s., in terms of weak convergence,

$$\mathbb{P}^{(k,\ell)}_{\infty,Y_{-k}^{-1}} \underset{\ell\to\infty}{\to} \mathbb{P}_{Y_0\,|\,Y_{-k}^{-1}}$$

(see for instance Theorem 3.1 in [18]). Next, with a slight modification of techniques of Theorem 2.2 in [14],

$$\int y^2 \mathrm{d}\mathbb{P}^{(k,\ell)}_{\infty,Y_{-k}^{-1}}(y) \underset{\ell\to\infty}{\to} \int y^2 \mathrm{d}\mathbb{P}_{Y_0\,|\,Y_{-k}^{-1}}(y) \quad \text{a.s.},$$

which leads to

$$\sup_{\ell\geq 0}\int y^2 \mathrm{d}\mathbb{P}^{(k,\ell)}_{\infty,Y_{-k}^{-1}}(y) < \infty \quad \text{a.s.}$$

Moreover, by assumption $(H3)$, the distribution function of $\mathbb{P}_{Y_0\,|\,Y_{-k}^{-1}}$ is a.s. increasing. Thus, setting

$$\left\{q_{\tau,\infty}^{(k,\ell)}(Y_{-k}^{-1})\right\} = \mathcal{Q}_\tau(\mathbb{P}^{(k,\ell)}_{\infty,Y_{-k}^{-1}}) \quad \text{and} \left\{q_\tau^{(k)}(Y_{-k}^{-1})\right\} = \mathcal{Q}_\tau(\mathbb{P}_{Y_0\,|\,Y_{-k}^{-1}})$$

and applying Lemma 5.4 yields

$$q_{\tau,\infty}^{(k,\ell)}(Y_{-k}^{-1}) \underset{\ell\to\infty}{\to} q_\tau^{(k)}(Y_{-k}^{-1}) \quad \text{a.s.}$$

Consequently,

$$\rho_\tau\left(Y_0 - T_\ell\left(q_{\tau,\infty}^{(k,\ell)}(Y_{-k}^{-1})\right)\right) \underset{\ell\to\infty}{\to} \rho_\tau\left(Y_0 - q_\tau^{(k)}(Y_{-k}^{-1})\right) \quad \text{a.s.}$$

It turns out that the above convergence also holds in mean. To see this, note first that

$$\begin{aligned}
&\rho_\tau\left(Y_0 - T_\ell\left(q_{\tau,\infty}^{(k,\ell)}(Y_{-k}^{-1})\right)\right)\\
&\quad= \rho_\tau\left(Y_0 - T_\ell(Y_0) + T_\ell(Y_0) - T_\ell\left(q_{\tau,\infty}^{(k,\ell)}(Y_{-k}^{-1})\right)\right)\\
&\quad\leq \rho_\tau\left(Y_0 - T_\ell(Y_0)\right) + \rho_\tau\left(T_\ell(Y_0) - T_\ell\left(q_{\tau,\infty}^{(k,\ell)}(Y_{-k}^{-1})\right)\right)\\
&\qquad\text{(by statement 2. of Lemma 5.3)}\\
&\quad\leq 2|Y_0| + \rho_\tau\left(Y_0 - q_{\tau,\infty}^{(k,\ell)}(Y_{-k}^{-1})\right) \quad \text{a.s.}\\
&\qquad\text{(by statement 3. of Lemma 5.3).}
\end{aligned}$$

Thus

$$\begin{aligned}
&\mathbb{E}\left[\left(\rho_\tau\left(Y_0 - T_\ell\left(q_{\tau,\infty}^{(k,\ell)}(Y_{-k}^{-1})\right)\right)\right)^2\right] \\
&\quad \leq \mathbb{E}\left[\left(2|Y_0| + \rho_\tau\left(Y_0 - q_{\tau,\infty}^{(k,\ell)}(Y_{-k}^{-1})\right)\right)^2\right] \\
&\quad \leq 8\mathbb{E}\left[Y_0^2\right] + 2\mathbb{E}\left[\left(\rho_\tau\left(Y_0 - q_{\tau,\infty}^{(k,\ell)}(Y_{-k}^{-1})\right)\right)^2\right].
\end{aligned}$$

In addition,

$$\begin{aligned}
&\sup_{\ell\geq 1}\mathbb{E}\left[\left(\rho_\tau\left(Y_0 - q_{\tau,\infty}^{(k,\ell)}(Y_{-k}^{-1})\right)\right)^2\right] \\
&\quad = \sup_{\ell\geq 1}\mathbb{E}\left[\left(\min_{q(.)}\mathbb{E}_{\mathbb{P}^{(k,\ell)}_{\infty,Y_{-k}^{-1}}}\rho_\tau\left(Y_0 - q(Y_{-k}^{-1})\right)\right)^2\right] \\
&\quad \leq \mathbb{E}\left[\left(\rho_\tau(Y_0)\right)^2\right] \\
&\qquad \text{(by Jensen's inequality)} \\
&\quad \leq \mathbb{E}\left[Y_0^2\right] < \infty.
\end{aligned}$$

This implies

$$\mathbb{E}\left[\left(\rho_\tau\left(Y_0 - T_\ell\left(q_{\tau,\infty}^{(k,\ell)}(Y_{-k}^{-1})\right)\right)\right)^2\right] < \infty,$$

i.e., the sequence is uniformly integrable. Thus we obtain, as desired,

$$\lim_{\ell\to\infty}\mathbb{E}\left[\rho_\tau\left(Y_0 - T_\ell\left(q_{\tau,\infty}^{(k,\ell)}(Y_{-k}^{-1})\right)\right)\right] = \mathbb{E}\left[\rho_\tau\left(Y_0 - q_\tau^{(k)}(Y_{-k}^{-1})\right)\right].$$

Putting all pieces together,

$$\begin{aligned}
\lim_{\ell\to\infty}\varepsilon_{k,\ell} &= \lim_{\ell\to\infty}\mathbb{E}\left[\rho_\tau\left(Y_0 - T_\ell\left(q_{\tau,\infty}^{(k,\ell)}(Y_{-k}^{-1})\right)\right)\right] \\
&= \mathbb{E}\left[\rho_\tau\left(Y_0 - q_\tau^{(k)}(Y_{-k}^{-1})\right)\right] \\
&\triangleq \varepsilon_k^\star.
\end{aligned}$$

It remains to prove that $\lim_{k\to\infty}\varepsilon_k^\star = L^\star$. To this aim, for all $k \geq 1$, let Z_k be the $\sigma(Y_{-k}^{-1})$-measurable random variable defined by

$$Z_k = \rho_\tau\left(Y_0 - q_\tau^{(k)}(Y_{-k}^{-1})\right) = \min_{q(.)}\mathbb{E}_{\mathbb{P}_{Y_0\,|\,Y_{-k}^{-1}}}\left[\rho_\tau\left(Y_0 - q(Y_{-k}^{-1})\right)\right].$$

Observe that $\{Z_k\}_{k=0}^{\infty}$ is a nonnegative supermartingale with respect to the family of sigma algebras $\{\sigma(Y_{-k}^{-1})\}_{k=1}^{\infty}$. In addition,

$$\begin{aligned}
\sup_{k\geq 1}\mathbb{E}[Z_k^2] &= \sup_{k\geq 1}\mathbb{E}\left[\left(\min_{q(.)}\mathbb{E}_{\mathbb{P}_{Y_0\,|\,Y_{-k}^{-1}}}\rho_\tau\left(Y_0 - q(Y_k^{-1})\right)\right)^2\right] \\
&\leq \sup_{k\geq 1}\mathbb{E}\left[\left(\mathbb{E}_{\mathbb{P}_{Y_0\,|\,Y_{-k}^{-1}}}\rho_\tau\left(Y_0\right)\right)^2\right] \\
&\leq \sup_{k\geq 1}\mathbb{E}\left[\left(\rho_\tau(Y_0)\right)^2\right] \\
&\quad \text{(by Jensen's inequality)} \\
&\leq \sup_{k\geq 1}\mathbb{E}[Y_0^2] < \infty.
\end{aligned}$$

Therefore,

$$\mathbb{E}[Z_k] \underset{k\to\infty}{\to} \mathbb{E}[Z_\infty],$$

where

$$Z_\infty = \min_{q(.)}\mathbb{E}_{\mathbb{P}_{Y_0\,|\,Y_{-\infty}^{-1}}}\left[\rho_\tau\left(Y_0 - q(Y_{-\infty}^{-1})\right)\right].$$

Consequently,

$$\lim_{k\to\infty}\varepsilon_k^\star = L^\star.$$

We finish the proof by using Lemma 5.2. On the one hand, a.s.,

$$\begin{aligned}
&\limsup_{n\to\infty}\inf_{k,\ell}\left(L_n\left(h_n^{(k,\ell)} - \frac{2\ln b_{k,\ell}}{n\eta_{n+1}}\right)\right) \\
&\leq \inf_{k,\ell}\left(\limsup_{n\to\infty} L_n\left(h_n^{(k,\ell)} - \frac{2\ln b_{k,\ell}}{n\eta_{n+1}}\right)\right) \\
&\leq \inf_{k,\ell}\left(\limsup_{n\to\infty} L_n\left(h_n^{(k,\ell)}\right)\right) \\
&= \inf_{k,\ell}\varepsilon_{k,\ell} \\
&\leq \lim_{k\to\infty}\lim_{\ell\to\infty}\varepsilon_{k,\ell} \\
&\leq L^\star.
\end{aligned}$$

Moreover,

$$\begin{aligned}
&\frac{1}{2n}\sum_{t=1}^{n}\eta_t\sum_{k,\ell=1}^{\infty}p_{k,\ell,n}\left[\rho_\tau\left(Y_t - h_t^{(k,\ell)}(Y_1^{t-1})\right)\right]^2 \\
&\leq \frac{1}{2n}\sum_{t=1}^{n}\eta_t\sum_{k=1}^{\infty}\sum_{\ell=1}^{\infty}p_{k,\ell,n}\left[\rho_\tau\left(Y_t - T_{\min(t^\delta,\ell)}\left(\bar{h}_t^{(k,\ell)}(Y_1^{t-1})\right)\right)\right]^2 \\
&\leq \frac{1}{2n}\sum_{t=1}^{n}\eta_t\sum_{k=1}^{\infty}\sum_{\ell=1}^{\infty}p_{k,\ell,n}\left|Y_t - T_{\min(t^\delta,\ell)}\left(\bar{h}_t^{(k,\ell)}(Y_1^{t-1})\right)\right|^2.
\end{aligned}$$

Thus

$$\begin{aligned}
&\frac{1}{2n}\sum_{t=1}^{n}\eta_t\sum_{k,\ell=1}^{\infty}p_{k,\ell,n}\left[\rho_\tau\left(Y_t-h_t^{(k,\ell)}(Y_1^{t-1})\right)\right]^2\\
&\quad\le\frac{1}{n}\sum_{t=1}^{n}\eta_t\sum_{k=1}^{\infty}\left(\sum_{\ell=1}^{\infty}p_{k,\ell,n}|Y_t|^2+\sum_{\ell=1}^{\infty}p_{k,\ell,n}\left[T_{\min(t^\delta,\ell)}\left(\bar{h}_t^{(k,\ell)}(Y_1^{t-1})\right)\right]^2\right)\\
&\quad\le\frac{1}{n}\sum_{t=1}^{n}\eta_t\sum_{k=1}^{\infty}\left(\sum_{\ell=1}^{\infty}p_{k,\ell,n}|Y_t|^2+\sum_{\ell=1}^{\infty}p_{k,\ell,n}t^{2\delta}\right)\\
&\quad\le\frac{1}{n}\sum_{t=1}^{n}\eta_t\sum_{k,\ell=1}^{\infty}p_{k,\ell,n}(Y_t^2+t^{2\delta})\\
&\quad=\frac{1}{n}\sum_{t=1}^{n}\eta_t(t^{2\delta}+Y_t^2).
\end{aligned}$$

Therefore, since $n^{2\delta}\eta_n\to 0$ as $n\to\infty$ and $\mathbb{E}[Y_0^2]<\infty$,

$$\limsup_{n\to\infty}\frac{1}{2n}\sum_{t=1}^{n}\eta_t\sum_{k,\ell=1}^{\infty}p_{k,\ell,n}\left[\rho_\tau\left(Y_t-h_t^{(k,\ell)}(Y_1^{t-1})\right)\right]^2=0\quad\text{a.s.}$$

Putting all pieces together, we obtain, a.s.,

$$\limsup_{n\to\infty}L_n(g)\le L^\star,$$

and this proves the result.

□

5.2 Proof of Lemma 2.1

To prove the first statement of the lemma, it will be enough to show that, for all $q\in\mathbb{R}$,

$$\mathbb{E}\left[\rho_\tau\left(Y-q\right)\right]-\mathbb{E}\left[\rho_\tau\left(Y-q_\tau\right)\right]\ge 0.$$

We separate the cases $q\ge q_\tau$ and $q<q_\tau$.

(*i*) If $q\ge q_\tau$, then

$$\begin{aligned}
&\mathbb{E}\left[\rho_\tau(Y-q)\right]-\mathbb{E}\left[\rho_\tau(Y-q_\tau)\right]\\
&\quad=\mathbb{E}\left[(Y-q)(\tau-\mathbf{1}_{[Y\le q]})-(Y-q_\tau)(\tau-\mathbf{1}_{[Y\le q_\tau]})\right]\\
&\quad=\mathbb{E}\left[(Y-q)\left(\tau-(\mathbf{1}_{[Y\le q_\tau]}+\mathbf{1}_{[q_\tau<Y\le q]})\right)-(Y-q_\tau)(\tau-\mathbf{1}_{[Y\le q_\tau]})\right]\\
&\quad=\mathbb{E}\left[(q_\tau-q)(\tau-\mathbf{1}_{[Y\le q_\tau]})\right]-\mathbb{E}\left[(Y-q)\mathbf{1}_{[q_\tau<Y\le q]}\right].
\end{aligned}$$

We have

$$\begin{aligned}\mathbb{E}\left[(q_\tau - q)(\tau - \mathbf{1}_{[Y \le q_\tau]})\right] &= (q_\tau - q)\left(\tau - \mathbb{P}[Y \le q_\tau]\right) \\ &= (q_\tau - q)\left[\tau - F_Y\left(F_Y^{\leftarrow}(\tau)\right)\right] \\ &\ge 0\end{aligned}$$

and, clearly,

$$-\mathbb{E}\left[(Y - q)\mathbf{1}_{[q_\tau < Y \le q]}\right] \ge 0.$$

This proves the desired statement.

(*ii*) If $q < q_\tau$, then

$$\begin{aligned}&\mathbb{E}\left[\rho_\tau\left(Y - q\right)\right] - \mathbb{E}\left[\rho_\tau(Y - q_\tau)\right] \\ &\quad = \mathbb{E}\left[(Y - q)(\tau - \mathbf{1}_{[Y \le q]}) - (Y - q_\tau)(\tau - \mathbf{1}_{[Y \le q_\tau]})\right] \\ &\quad = \mathbb{E}\left[(Y - q)(\tau - \mathbf{1}_{[Y \le q]}) - (Y - q_\tau)\left(\tau - (\mathbf{1}_{[Y \le q]} + \mathbf{1}_{[q < Y \le q_\tau]})\right)\right] \\ &\quad = \mathbb{E}\left[(q_\tau - q)(\tau - \mathbf{1}_{[Y \le q]})\right] - \mathbb{E}\left[(Y - q_\tau)(\tau - \mathbf{1}_{[q < Y \le q_\tau]})\right].\end{aligned}$$

For $q < q_\tau$, $\mathbb{P}\left[Y \le q\right] = F_Y(q) < \tau$. Consequently

$$\mathbb{E}\left[(q_\tau - q)(\tau - \mathbf{1}_{[Y \le q]})\right] > 0.$$

Since

$$-\mathbb{E}\left[(Y - q_\tau)(\tau - \mathbf{1}_{[q < Y \le q_\tau]})\right] \ge 0,$$

we are led to the desired result.

Suppose now that F_Y is increasing. To establish the second statement of the lemma, a quick inspection of the proof reveals that it is enough to prove that, for $q > q_\tau$,

$$\mathbb{E}\left[(Y - q)\mathbf{1}_{[q_\tau < Y \le q]}\right] < 0.$$

Take $q' \in (q_\tau, q)$ and set $S = [q_\tau < Y \le q']$. Clearly,

$$\mathbb{P}(S) = F_Y(q') - F_Y(q_\tau) > 0.$$

Therefore

$$\mathbb{E}\left[(Y - q)\mathbf{1}_{[q_\tau < Y \le q]}\right] \le \mathbb{E}[(Y - q)\mathbf{1}_S] < 0.$$

□

Acknowledgments. The authors are greatly indebted to Adrien Saumard and Joannès Vermorel for their valuable comments and insightful suggestions on the first draft of the paper. They also thank two referees and the Associate Editor for their careful reading of the paper and thoughtful critical comments.

References

[1] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, 2nd ed. New York: Springer-Verlag, 1991.

[2] L. Györfi, W. Härdle, P. Sarda, and P. Vieu, *Nonparametric Curve Estimation from Time Series*. Berlin: Springer-Verlag, 1989.

[3] D. Bosq, *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*. New York: Springer-Verlag, 1996.

[4] A. Gannoun, J. Saracco, and K. Yu, "Nonparametric prediction by conditional median and quantiles," *J. Statist. Plann. Inference*, vol. 117, pp. 207–223, 2003.

[5] R. Koenker and K. F. Hallock, "Quantile regression," *J. Eco. Persp.*, vol. 15, pp. 143–156, 2001.

[6] D. Duffie and J. Pan, "An overview of value at risk," *J. Derivatives*, vol. 4, pp. 7–49, 1997.

[7] R. Koenker, *Quantile Regression*. Cambridge: Cambridge University Press, 2005.

[8] N. Cesa-Bianchi and G. Lugosi, *Prediction, Learning, and Games*. New York: Cambridge University Press, 2006.

[9] R. Koenker and G. Bassett, Jr., "Regression quantiles," *Econometrica*, vol. 46, pp. 33–50, 1978.

[10] P. Algoet, "The strong law of large numbers for sequential decisions under uncertainty," *IEEE Trans. Inform. Theory*, vol. 40, pp. 609–633, 1994.

[11] L. Györfi and G. Lugosi, "Strategies for sequential prediction of stationary time series," in *Modeling Uncertainty*, ser. Internat. Ser. Oper. Res. Management Sci. Boston: Kluwer Acad. Publ., 2002, vol. 46, pp. 225–248.

[12] A. B. Nobel, "On optimal sequential prediction for general processes," *IEEE Trans. Inform. Theory*, vol. 49, pp. 83–98, 2003.

[13] L. Györfi and G. Ottucsák, "Sequential prediction of unbounded sationary time series," *IEEE Trans. Inform. Theory*, vol. 53, pp. 1866–1872, 2007.

[14] G. Biau, K. Bleakley, L. Györfi, and G. Ottucsák, "Nonparametric sequential prediction of time series," *J. Nonparametr. Stat.*, vol. 22, pp. 297–317, 2010.

[15] P. Algoet, "Universal schemes for prediction, gambling and portfolio selection," *Ann. Probab.*, vol. 20, pp. 901–941, 1992.

[16] L. Györfi and D. Schäfer, "Nonparametric prediction," *Advances in Learning Theory: Methods, Models and Applications*, pp. 341–356, 2003.

[17] L. Györfi, G. Lugosi, and F. Udina, "Nonparametric kernel-based sequential investment strategies," *Math. Finance*, vol. 16, pp. 337–357, 2006.

[18] L. Györfi, F. Udina, and H. Walk, "Nonparametric nearest neighbor based empirical portfolio selection strategies," *Statist. Decisions*, vol. 26, pp. 145–157, 2008.

[19] L. Györfi, M. Kohler, A. Krzyżak, and H. Walk, *A Distribution-Free Theory of Nonparametric Regression*. New York: Springer-Verlag, 2002.

[20] F. Bunea and A. Nobel, "Sequential procedures for aggregating arbitrary estimators of a conditional mean," *IEEE Trans. Inform. Theory*, vol. 54, pp. 1725–1735, 2008.

[21] L. Györfi, F. Udina, and H. Walk, "Experiments on universal portfolio selection using data from real markets," 2008, technical report. [Online]. Available: http://tukey.upf.es/papers/NNexp.pdf

[22] J. O. Street, R. J. Carroll, and D. Ruppert, "A note on computing robust regression estimates via iteratively reweighted least squares," *Amer. Statistician*, vol. 42, pp. 152–154, 1988.

[23] I. Takeuchi, Q. V. Le, T. D. Sears, and A. J. Smola, "Nonparametric quantile estimation," *J. Mach. Learn. Res.*, vol. 7, pp. 1231–1264, 2006.

[24] P. Hall, L. Peng, and Q. Yao, "Prediction and nonparametric estimation for time series with heavy tails," *J. Time Ser. Anal.*, vol. 23, pp. 313–331, 2002.

[25] S. G. Madrikakis, S. C. Wheelwright, and R. J. Hyndman, *Forecasting, Methods and Applications*. New York: Wiley, 1998.

[26] L. Breiman, "The individual ergodic theorem of information theory," *Ann. Math. Statist.*, vol. 28, pp. 809–811, 1957.

[27] P. Billingsley, *Probability and Measure*, 3rd ed. New York: John Wiley & Sons, 1995.